

USITP 94-11

July 1994

The QCD Trace Anomaly as a Vacuum Effect

(The vacuum is a medium is the message!)

J. Grundberg^{*} and T. H. Hansson[†]

We use arguments taken from the electrodynamics of media to deduce the QCD trace anomaly from the expression for the vacuum energy in the presence of an external color magnetic field.

Submitted to: *American Journal of Physics*

^{*} Department of Mathematics and Physics, Mälardalens Högskola
Box 11, S-72103 Västerås, Sweden
730 Sweden

[†] Supported by the Swedish Natural Science Research Council

I. Introduction

Intuitive discussions of phenomena in quantum field theory frequently take the form of describing them as properties of the vacuum. The vacuum, it is then argued, is just like any other medium, so we can use the intuition developed in other fields of physics to get an understanding of quantum field theory. Perhaps the first example of this was the interpretation of unoccupied negative energy states in Dirac's relativistic electron theory as positrons. Today the obvious analogy is the behavior of electrons in semiconductors, though historically it seems that the concept of holes developed independently in quantum field theory and solid state physics [1]. Another example is the Casimir effect, which is explained in terms of the zero-point fluctuations of the quantized electromagnetic field. A third example is spontaneous symmetry breaking of the vacuum, where the intuition is supplied by the behavior of (anti-)ferromagnets and (for local symmetries) superconductors. As a final example we have the charge-screening properties of the vacuum. Already in the 30s it was realised that the QED vacuum behaves just like an ordinary dielectric medium [2]: Charges placed in the vacuum get screened by polarization effects. Around 1980 Nielsen, [3], and Hughes, [4], showed that asymptotic freedom, the antiscreening property of the QCD vacuum, could be interpreted as (color-) paramagnetism.¹ In this paper we will show that the calculation this result rests upon, the change in vacuum energy due to an external (abelian color-) magnetic field, can also be used to derive the QCD trace anomaly. The point is that the energy is all we need in order to apply some elementary arguments borrowed from the electrodynamics of media to get the stress tensor, and hence the anomaly. The vacuum is indeed just another (color-) electromagnetic medium.

Our paper is organized as follows. We first (section 2) review the medium picture of screening (QED) and antiscreening (QCD). In the course of this discussion we quote the result for the vacuum energy in QCD and introduce the renormalization group equation for charges. In section 3 we then use the expression for the vacuum energy and some "elementary physics" arguments to deduce the stress tensor for QCD and calculate the anomaly. In these two sections we have tried to keep the discussion at an elementary level and to explain those results that we have just quoted from the literature. Finally, in section 4, we conclude with some comments as to how our derivation of the trace anomaly reflects the breaking of scale invariance. In appendix A we give a standard derivation of the trace anomaly and compare some details with our derivation. We also briefly discuss higher order effects and some

¹Apparently this had earlier been realised by 't Hooft, but not published by him [5].

other technicalities. In appendix B we justify an argument based on Lorentz invariance used in the text. Both appendices requires more knowledge of quantum field theory than the main text.

Although this work is a rederivation of a known result, we hope that it has a value beyond the purely pedagogical. The vacuum as a medium has been a fruitful picture in trying to develop intuition about non Abelian gauge theories. In order to make good use of the medium picture, and avoid pitfalls due to false analogies, we think it is of value to establish its limits as precisely as possible.

2. Charge (anti)screening

As a preliminary to the discussion of the trace anomaly we wish to remind the reader about the intuitive interpretation of charge (anti-)screening. Starting with QED, we consider two heavy particles of charge e , separated by a distance r . Classically the potential energy of the pair is given by

$$E(r) = \frac{\alpha}{r} \quad , \quad (1)$$

where $\alpha = e^2/4\pi$. This result is modified in QED. To $O(\alpha^2)$ the potential energy is instead given by

$$E(r) = \frac{1}{\epsilon(r)} \frac{\alpha}{r} \quad , \quad (2)$$

with [6]

$$\frac{1}{\epsilon(r)} = 1 - \frac{\alpha}{3\pi} \ln \frac{\Omega^2}{m^2} + \frac{2\alpha}{3\pi} \int_m^\infty du e^{-2ru} \left[\frac{2u^2 + m^2}{2u^2} \right] \frac{(u^2 - m^2)^{1/2}}{u^2} \quad , \quad (3)$$

where we have introduced an ultraviolet cut-off Ω . The usual charge renormalization in QED is obtained by taking

$$\alpha_{ren} = \alpha \left(1 - \frac{\alpha}{3\pi} \ln \frac{\Omega^2}{m^2} \right) \quad (4)$$

as the physical charge. This amounts to removing the cut-off dependent term from equation (3), and in the limit $r \rightarrow \infty$, we get

$$\frac{1}{\epsilon(r)} \rightarrow 1 + \frac{\alpha}{4\sqrt{\pi}(mr)^{3/2}} e^{-2mr} . \quad (5)$$

The "effective charge" defined by

$$E(r) = \frac{\alpha_{eff}}{r} \quad , \quad (6)$$

depends on the distance. The intuitive explanation of this is that the QED vacuum behaves like a polarizable medium. In Dirac's original picture the vacuum has a homogeneous distribution of negative energy electrons. When an external charge is introduced, the charge distribution is modified. More negative charge accumulates around a positive external charge, less around a negative external charge. The effect is that the external charge is screened. This way of looking at the problem is not explicitly charge-symmetric, but the same intuitive picture works if we prefer to view the vacuum as consisting of (virtual) electron-positron pairs.

The quantitative result for $\epsilon(r)$ given in (5) has the property that $\epsilon \rightarrow 1$ as $r \rightarrow \infty$. This limit is finite because the electron mass $m \neq 0$ and we have used this to renormalize the electric charge in such a way that α means the charge at the "preferred" distance $r = \infty$, but this is clearly a matter of choice. A situation more like QCD arises when $m = 0$. In the limit $m \ll 1/r$, we get

$$\delta E(r) = \frac{\alpha}{r} \left[-\frac{\alpha}{3\pi} \ln \frac{\Omega^2}{m^2} + \frac{2\alpha}{3\pi} \int_m^\infty du \frac{e^{-2ru}}{u} \right] \quad . \quad (7)$$

Note that both terms are logarithmically divergent as $m \rightarrow 0$, but that the sum is finite. The charge renormalization can now be done in the following way: First, note that in the limit $m \rightarrow 0$,

$$r\delta E(r) - r_0\delta E(r_0) = -\frac{2\alpha^2}{3\pi} \ln \frac{r}{r_0} \quad , \quad (8)$$

so we can write the total potential energy as,

$$\begin{aligned} E &= \frac{\alpha}{r} + \delta E(r) \\ &= \frac{\alpha + r_0\delta E(r_0)}{r} - \frac{2\alpha^2}{3\pi r} \ln \frac{r}{r_0} \quad . \end{aligned} \quad (9)$$

Then we define

$$\alpha(r_0) = \alpha + r_0\delta E(r_0) \quad , \quad (10)$$

to get, to $O(\alpha^2)$,

$$E(r) = \frac{\alpha(r_0)}{r} \left(1 - \frac{2\alpha(r_0)}{3\pi} \ln \frac{r}{r_0} \right) \quad . \quad (11)$$

It is no longer possible to take the limit $r_0 \rightarrow \infty$ in defining the charge, instead we pick an arbitrary scale r_0 . The dependence of the charge parameter on

the defining distance is governed by the statement that the energy $E(r)$ is a physical quantity and should be independent of arbitrary choices like r_0 . Thus

$$\frac{dE}{dr_0} = 0 \quad , \quad (12)$$

which, to this order in α , gives

$$r_0 \frac{d\alpha}{dr_0} = -\frac{2\alpha^2}{3\pi} \quad . \quad (13)$$

This is the renormalization group equation for α . It is conventional to write renormalization group equations in terms of derivatives with respect to a mass scale μ rather than a length scale. This merely reverses the sign of the right hand side of the equation and we get

$$\mu \frac{d\alpha}{d\mu} = \beta(\alpha) = \frac{2\alpha^2}{3\pi} \quad . \quad (14)$$

The description of screening, as formulated above, is in terms of electric fields from a point charge. One can also discuss the effect in terms of the response of the vacuum to an external homogeneous electromagnetic field. Since an external electric field always gives a non-vanishing probability for pair creation, it is more convenient to discuss the response to an external magnetic field and calculate the magnetic permeability μ (not to be confused with the mass scale in the renormalization group equation!) rather than ϵ . We can then deduce ϵ from μ if we assume that the vacuum is Lorentz-invariant. If that is the case (for further discussion, see Appendix B), then we should have

$$\epsilon\mu = 1 \quad , \quad (15)$$

so, if $\mu > 1$ (paramagnetic vacuum), then $\epsilon < 1$ (antiscreening) and if $\mu < 1$ (diamagnetic vacuum), then $\epsilon > 1$ (screening). The strategy is thus to calculate the energy of the vacuum in an external magnetic field and read off μ . This was done in [3], [4], for the case of a constant homogeneous color magnetic field belonging to an abelian subgroup of the color group, *i.e.* of the type

$$\mathbf{A}^a(x) = \mathbf{A}(x)T^a \quad (16)$$

$$\mathbf{B}^a(x) = \nabla \times \mathbf{A}(x)T^a \quad (17)$$

where T^a is a constant color matrix which is usually taken to be diagonal (*i.e.* for SU(2) proportional to σ^3 and for SU(3) *e.g.* proportional to λ^8). We refer

the reader to the above references for the details, here we will just outline the calculation and give the final result.

The vacuum energy is calculated by summing the zero-point energies for the modes of all fields. For the gluon field this is $\frac{\hbar\omega}{2}$ for each mode, the usual harmonic oscillator zero-point energy, for the quarks the exclusion principle (technically, the fact that fermion creation/annihilation operators satisfy anticommutation relations instead of commutation relations) reverses the sign of this, so the contribution is $-\frac{\hbar\omega}{2}$ for each mode, counting particles and antiparticles separately. In the Dirac sea picture, one can also think of the fermionic part as coming from the filled negative energy states, where each mode contributes $-\hbar\omega$.

The frequencies for the various modes are essentially the energy eigenvalues for particles with spin 1 (gluons) or spin 1/2 (quarks) in an external magnetic field. As is well known from elementary solid state physics (in the case of spin 1/2), both the spin and the orbital magnetic moment give contributions. Remember that without spin, an ordinary electron gas would be diamagnetic, with the spin included it is actually paramagnetic. In the case of the vacuum medium the situation is reversed. The vacuum electrons give a diamagnetic contribution due to the extra sign coming from the exclusion principle. For gluons the story is similar, but since there is no extra minus sign, their contribution is paramagnetic. Of course, when we sum over all modes, the energy actually diverges. One has to introduce a cut-off Ω and subtract a piece $\propto \Omega^4$ that is independent of the external field. The result for the field-dependent piece in QCD with N_F species of mass-less colored fermions is

$$U_{vac} = \frac{V}{2} \left[B^2 + (gB)^2 \frac{33 - 2N_F}{48\pi^2} \ln \frac{\Omega^2}{|gB|} \right] \quad , \quad (18)$$

where g is the QCD coupling constant and V is the volume of the box whose energy we have calculated. The first term in (18) is the classical contribution to the energy, just like the Coulomb contribution in the case of external point charges. We still have an explicit Ω in this expression. We can handle it in a way that is similar to what we did in the point charge case above and which allows us to let $\Omega \rightarrow \infty$. By rearranging the expression in (18) we can write it as follows:

$$u_{vac} = \frac{U_{vac}}{V} = \frac{1}{2} B^2 \left[1 + g^2 \frac{33 - 2N_F}{48\pi^2} \ln \frac{\Omega^2}{|gB_0|} \right] + \frac{1}{2} (gB)^2 \frac{33 - 2N_F}{48\pi^2} \ln \left| \frac{gB_0}{gB} \right| \quad . \quad (19)$$

We now renormalize the field strength, in analogy with (10), by defining

$$B_{ren}(gB_0) = B \left[1 + g^2 \frac{33 - 2N_F}{96\pi^2} \ln \frac{\Omega^2}{|gB_0|} \right] , \quad (20)$$

so, to $O(g^2)$,

$$u_{vac} = \frac{B(gB_0)^2}{2} \left[1 + g^2 \frac{33 - 2N_F}{48\pi^2} \ln \left| \frac{gB(gB_0)}{gB_0} \right| \right] , \quad (21)$$

where all field strengths are renormalized ones. Just as we did for the charge in QED, we can now derive a renormalization group equation for $B(gB_0)$. The result is

$$gB_0 \frac{dB}{d(gB_0)} = \frac{33 - 2N_F}{96\pi^2} g^2 . \quad (22)$$

If we remember that gB is a physical quantity, we can readily convert this to a renormalization group equation for g : $g_r B_r = gB$, so

$$gB_0 \frac{dg_r}{d(gB_0)} = -\frac{33 - 2N_F}{96\pi^2} g^3 = \frac{\beta(g_r)}{2} , \quad (23)$$

where the coupling constants are renormalized ones. We have introduced the conventional β -function, and the factor $1/2$ in the last equality arises because gB_0 has dimension $(\text{mass})^2$, and, as we said above, it is conventional to use the derivative with respect to something with dimension mass.

To deduce a permeability from the vacuum energy, one usually defines μ by

$$u = \frac{1}{2} H B = \frac{1}{2\mu} B^2 , \quad (24)$$

where the first equation has been taken from the electrodynamics of linear media. This is strictly speaking not quite correct, and we will return to the relation between H , B and u in the next section, but for the present purposes it suffices. We get

$$\mu = 1 + \frac{33 - 2N_F}{48\pi^2} \ln \left| \frac{gB_0}{gB} \right| , \quad (25)$$

so we see that if $N_F < 16$, $\mu > 1$, the QCD vacuum is paramagnetic, and, by the previous argument $\epsilon < 1$, the QCD vacuum antiscreens color charge.

3. The trace anomaly

The point of view we wish to emphasize in this paper is that the QCD vacuum is like any other nonlinear medium. In particular, once we have the energy as a function of the appropriate variables, we can use elementary virtual work arguments to deduce the stress tensor and thus calculate the trace anomaly. However, before turning to that argument we shall ask what the medium we are talking about really is.

In ordinary electrodynamics for media, the inhomogeneous Maxwells equations are given in terms of the fields \vec{D} and \vec{H} and the corresponding macroscopic currents and charges. The induced current and charge distributions in the medium is taken into account via polarization and magnetization which relates the \vec{D} and \vec{H} fields to the fields \vec{E} and \vec{B} . Since the fields and the macroscopic currents are related by Maxwells equations, we can choose to either consider the currents or the fields as fixed. When discussing the vacuum the situation is less clear, since there is no obvious separation between external (or macroscopic) and internal (or microscopic) currents. Roughly speaking, it is the virtual particles that constitute the medium, but we have to make this statement a bit more precise. In QED the situation is rather simple if we consider the case of an external field and work only to lowest order in the coupling constant. In this case the only virtual particles are electron - positron pairs, and the calculation is exactly the one given in the previous section. Since there are no virtual photons, there is no ambiguity in how to define the background field. Note that if we had chosen to start from external currents of electrons rather than from a background field, we would have had to face the problem how to distinguish the "real" (on shell) electrons in the currents from the virtual (off shell) electrons in the vacuum medium.

In the case of QCD the situation is more subtle. Here there are virtual gluons produced to lowest order, and these have to be distinguished from the gluons that constitute the external field. This is done by employing the so called background field method. (For a review see *e.g.* [7]). This allows for a separation of the gluon field in a background and a quantum part. Only the quantum part appears in loops and needs to be gauge fixed. The background (or external) field is treated as classical, but due to the quantum loops it gets a non-quadratic (but gauge-invariant) effective action. The separation between background and quantum part of the gluon field depends on the gauge-fixing and renormalization procedure and is thus to a certain extent arbitrary. Nevertheless it will provide a strict definition of macroscopic and microscopic and thus allows us to use results from the electrodynamics of media, just as in the case of QED.

In order to identify the background field, which was introduced in (17), and

that appears in the expression (21) for the vacuum energy, we first note that the field (17) which enters our expressions is the curl of the vector potential (16). Thus it is divergence free, and should be a \mathbf{B} rather than a \mathbf{H} -field. Second, our complete hamiltonian density, including the background field energy, is (for the time being we put a tilde on the field to remind us that our aim is to identify it)

$$\mathcal{H} = \frac{\tilde{B}^2}{2} + \mathcal{H}_{matter}(\mathbf{p} - g\mathbf{A}_{quant} - g\tilde{\mathbf{A}}) \quad , \quad (26)$$

where \mathcal{H}_{matter} is the hamiltonian density for quarks and (the quadratic part of the) hamiltonian for gluons in an external magnetic field. For a constant and homogeneous external field we can choose the vector potential $\tilde{\mathbf{A}} = \frac{1}{2}\tilde{\mathbf{B}} \times \mathbf{r}$. Varying the field we get

$$\delta u_{vac} = \delta \langle vac | \mathcal{H} | vac \rangle = (\tilde{\mathbf{B}} - \mathbf{M}) \cdot \delta \tilde{\mathbf{B}} \quad , \quad (27)$$

where we have identified the magnetization,

$$\mathbf{M} = \langle vac | -\frac{1}{2}\mathbf{r} \times \frac{\partial \mathcal{H}_{matter}}{\partial \tilde{\mathbf{A}}} | vac \rangle = \langle vac | \frac{1}{2}\mathbf{r} \times \mathbf{j} | vac \rangle \quad , \quad (28)$$

where \mathbf{j} is the current operator. In the non relativistic case the right hand side in (28) reduces to a sum of the well know expressions for the orbital and spin magnetic moment densities. In the relativistic case, relevant for light quarks, there is no clean separation between orbital and spin contributions on the operator level, but the distinction can still be made by considering the following expression for the (unrenormalized) one loop vacuum energy[3]

$$u_{vac} = \frac{1}{2}B^2 - \frac{1}{2}B^2(-1)^{2a} \frac{\hat{g}^2}{4\pi^2} \sum_{m=-s}^s \left(m^2 - \frac{1}{12} \right) \ln \left| \frac{\Omega^2}{gB} \right| \quad , \quad (29)$$

which is valid for spin $s = 0, \frac{1}{2}, 1$, and where \hat{g} depends on the color representation of the particles. The term $\frac{1}{12}$ comes from the orbital motion and is common for both scalar, spinor and vector particles, while the term m^2 , where m is the component of the spin in the direction of the \mathbf{B} field, does depend on the spin. For gluons this expression coincides with (18) if $\hat{g}^2 = \frac{3}{2}g^2$, corresponding to adjoint gluons.

To identify \mathbf{H} and \mathbf{B} we compare δu_{vac} with the work done as the magnetic field is varied. As the field changes an electric field \mathbf{E} is induced and the work done on the system is [8],

$$\delta W = -\delta t \int d^3x \mathbf{E} \cdot \mathbf{j} = \int d^3x \delta \mathbf{A} \cdot \mathbf{j}$$

$$= \int d^3x \delta \mathbf{A} \cdot (\nabla \times \mathbf{H}) = \int d^3x \mathbf{H} \cdot \delta \mathbf{B} \quad , \quad (30)$$

where we have used Maxwells equation

$$\nabla \times \mathbf{H} = \frac{1}{c} \mathbf{j} \quad , \quad (31)$$

and \mathbf{j} is the current density that serves as external source for the magnetic field. The work per unit volume (30) should equal the change in energy (27):

$$\mathbf{H} \cdot \delta \mathbf{B} = (\tilde{\mathbf{B}} - \mathbf{M}) \cdot \delta \tilde{\mathbf{B}} \quad . \quad (32)$$

Since $\mathbf{B} = \mathbf{H} + \mathbf{M}$, we identify

$$\mathbf{B} = \tilde{\mathbf{B}} \quad , \quad (33)$$

and

$$\mathbf{H} = \frac{\partial u(B)}{\partial \mathbf{B}} \quad . \quad (34)$$

The work we discussed above is the total work on both medium and sources of the magnetic field. To be able to deduce the stress tensor we need to consider a situation where we can divide this work into work done on the medium and on the sources. This is what we are going to do next.

In order to discuss the work done in changing the magnetic field it is convenient to have a specific geometry in mind. We consider a cylindrical solenoid of length L , radius R and with N turns. A current I through the solenoid generates a \mathbf{H} -field, which, in the limit $L \rightarrow \infty$, N/L finite, is homogeneous, parallel to the axis of the solenoid and has magnitude

$$H = \frac{1}{c} \frac{N}{L} I \quad . \quad (35)$$

This solenoid encloses our medium, *i.e.*, the vacuum. We can define the stress tensor in terms of the virtual work done as we change the enclosed region while keeping the current constant. (The more familiar electrostatic analogue of this procedure is to consider a region bounded by capacitor plates and the work done as the plates are moved while keeping them at constant potential [9].) First, it is clear from the symmetry of the problem that the only non-vanishing components of the stress-tensor T_{ij} are T_{zz} (where the z-axis is the axis of the

solenoid) and $T_{xx} = T_{yy}$. Thus T_{ij} simply acts like an anisotropic pressure. As we change the solenoid we have to do a virtual work per unit volume

$$\delta w_1 = -p \frac{\delta V}{V} = -T_{zz} \frac{\delta L}{L} - T_{xx} \frac{\delta(\pi R^2)}{\pi R^2} \quad . \quad (36)$$

In addition to this work on the medium we also have to do some work to keep the current constant. As the solenoid is modified the magnetic flux changes and an emf,

$$\mathcal{E} = -\frac{1}{c} \frac{d}{dt} (B \pi R^2) \quad , \quad (37)$$

is induced. This emf does some work which has to be compensated for to keep the current constant. The additional work per unit volume is given by ([8], p 214)

$$\begin{aligned} \delta w_2 &= -\frac{NI}{\pi R^2 L} \mathcal{E} \delta t = \frac{1}{\pi R^2 L c} \frac{d}{dt} (\pi R^2 N I B) \delta t \\ &= \frac{1}{\pi R^2 L} \delta(\pi R^2 L H B) \quad , \end{aligned} \quad (38)$$

where we have used that I is constant and the relation between I and H . The total work per unit volume is thus

$$\begin{aligned} \delta w = \delta w_1 + \delta w_2 &= (HB - T_{zz}) \frac{\delta L}{L} + (HB - T_{xx}) \frac{2\delta R}{R} + \delta(HB) \\ &= (-H^2 \frac{dB}{dH} - T_{zz}) \frac{\delta L}{L} + (HB - T_{xx}) \frac{2\delta R}{R} \quad , \end{aligned} \quad (39)$$

where we have used that H , and consequently B , is independent of R and also that $H \propto L^{-1}$, *i.e.* $L\delta H = -H\delta L$. This work should equal the change in energy of the system per unit volume. This change has two terms. The first is due to that the solenoid encloses a different volume, the second is due to the changed magnetic field. The result is

$$\delta w = u \frac{\delta V}{V} + \delta u = u \frac{2\delta R}{R} + (u - H^2 \frac{dB}{dH}) \frac{\delta L}{L} \quad . \quad (40)$$

In the last step we used that $\delta u = H\delta B$, *i.e.* , equation (34), and again that $H \propto L^{-1}$. We can conclude that,

$$\begin{aligned} T_{zz} &= -u \\ T_{xx} &= T_{yy} = HB - u \quad , \end{aligned} \quad (41)$$

and hence the trace of the stress-energy tensor is

$$T^\mu{}_\mu = T_{00} - T_{xx} - T_{yy} - T_{zz} = 4u - 2HB \quad . \quad (42)$$

From section 2 we know that

$$u = \frac{B^2}{2} + \frac{33 - 2N_F}{48\pi^2} \frac{g^2 B^2}{2} \ln \left| \frac{gB}{gB_0} \right| \quad , \quad (43)$$

and thus the final result

$$T^\mu{}_\mu = -\frac{33 - 2N_F}{48\pi^2} g^2 B^2 = \frac{\beta(g)}{2g} F^{\mu\nu} F_{\mu\nu} \quad , \quad (44)$$

where we have used the lowest order result for the β -function,

$$\beta(g) = -\frac{33 - 2N_F}{48\pi^2} g^3 \quad . \quad (45)$$

We have thus recovered the usual result for the trace anomaly. Note that it was important in this derivation to have the correct relation between u , B and H . If we had applied linear-medium relations, the anomaly would have vanished. The entire effect comes from the non-linearity of the vacuum.

Readers wishing for a comparison of the T_{ij} we derived with a conventional field theory version are referred to Appendix A, where we also show that the coefficient in front of $F^{\mu\nu} F_{\mu\nu}$ is indeed related to the β -function.

4. The trace anomaly and scale braking

In the standard derivation of the trace anomaly, the relation to spontaneously broken scale invariance is emphasised. At a classical level the fields scale according to their canonical dimension,

$$\mathbf{B}(x) \rightarrow \lambda^{-2} \mathbf{B}(\lambda x) \quad (46)$$

$$\mathbf{E}(x) \rightarrow \lambda^{-2} \mathbf{E}(\lambda x) \quad (47)$$

under the rescaling $x^\mu \rightarrow \lambda x^\mu$ of the coordinates. Since the classical Lagrangian for the electromagnetic or gluon field does not contain any dimensionful parameter, this scale transformation is a symmetry and there is a corresponding conserved current, the so called dilatation (or scaling) current given by

$$j_{dil.}^\mu = T_\nu^\mu x^\nu \quad . \quad (48)$$

Since the energy-momentum tensor is conserved, we get

$$\partial_\mu j_{dil.}^\mu = T_\mu^\mu \quad , \quad (49)$$

so if the dilatation current is conserved, the trace of the energy-momentum tensor vanishes. (For a more careful discussion of this statement, see [10].) The trace anomaly thus implies that the dilatation current is not conserved. The classical scale invariance has been broken by quantum effects. How did this happen? A look at our calculations shows that we at some stage had to introduce a scale to be able to define the theory. In the case of external charges in QED we needed the length scale r_0 to define the charge, and in the case of an external homogeneous magnetostatic field we introduced the scale in the guise of a reference magnetic field gB_0 . In both cases we made the point that this reference scale (r_0 or gB_0) was arbitrary and we used this fact to derive renormalization group equations. However, despite the arbitrariness of the scale, it *had* to be introduced, and its presence destroys the scale invariance. As we change reference scale and coupling constant, there is a combination of them that remains unchanged: A renormalization group invariant scale that can be used to parametrize the theory. By integrating the renormalization group equation

$$\mu \frac{dg}{d\mu} = \beta(g) \quad , \quad (50)$$

we can trade the arbitrary scale μ for a (dimensionful) integration constant Λ

$$\Lambda = \mu e^{-\int_{g(\Lambda)}^{g(\mu)} \frac{dg}{\beta(g)}} \quad . \quad (51)$$

One might ask what is the advantage of having replaced the scale μ by the (seemingly) equally arbitrary scale Λ . This becomes clear if we reexpress the vacuum energy u_{vac} in (21) in terms of a Λ_B defined via (51) with $\mu^2 = gB_0$:

$$u_{vac} = -\frac{33 - 2N_F}{192\pi^2} \Lambda_B^4 \quad (52)$$

We see that Λ_B is directly related to a physical quantity, while μ is not. In fact, equation (51) describes a curve in the $g-\mu$ plane, and all points on this curve corresponds to the same physics (masses, scattering amplitudes *etc.*). To get different physics we have to change to a different curve, *i.e.* change the value of Λ . It is, however, important to realize that the actual numerical value (in *e.g.* MeV) of Λ depends on how the renormalization is performed. That is why we used a subscript B on the Λ in (52). There are many different prescriptions for defining Λ , but they can all be related, even though the required calculations sometimes can be lengthy.²

² For instance, to relate the Λ that occurs naturally in lattice calculations to the one used in perturbative QCD, one must perform a rather complicated perturbative lattice calculation.

Classically the theory is scale invariant and is characterized by a dimensionless parameter, g . Quantum mechanics breaks the scale invariance and transmutes g to a dimensionful parameter, Λ ! In QCD with mass less quarks this is the only parameter that sets the scale for hadron masses, crosssections, *etc.*, and it is used to parametrize theoretical predictions for QCD processes.

To summarize: The trace anomaly implies that scale invariance is broken by quantum corrections, and this occurs because, in one way or another, the renormalization procedure forces us to introduce a scale to define the theory.

Acknowledgement

We thank R.J. Hughes for a conversation about this and other possible ways of giving an elementary discussion of the trace anomaly, and Ken Johnson for helpful comments about the manuscript.

Appendix A.

In section 3 of this paper we used elementary arguments to deduce the stress-tensor from the energy density. In this appendix we first show that this way of deducing the stress-tensor coincides with the conventional field theory definition in terms of a variation of the effective action with respect to the metric. By combining the effective action method with renormalization group arguments we then show that the coefficient in the trace anomaly is indeed related to the β -function. (The argument here is the inverse of the one presented in [11]. We thank R.J. Hughes for informing us about this reference.)

The starting point for our field theory considerations is the renormalized effective action

$$\Gamma = \int d^4x \sqrt{-g} \mathcal{L} \quad . \quad (\text{A.1})$$

The stress-energy tensor is then given by³

$$T_{\alpha\beta} = \frac{2}{\sqrt{-g}} \frac{\delta \Gamma}{\delta g^{\alpha\beta}} = 2 \frac{\partial \mathcal{L}}{\partial g^{\alpha\beta}} - g_{\alpha\beta} \mathcal{L} \quad . \quad (\text{A.2})$$

We are interested in $T_{\alpha\beta}$ for a constant (abelian color) electromagnetic field, in which case $\mathcal{L} = \mathcal{L}(x, y)$ where $x = F^{\alpha\beta} F_{\alpha\beta}$ and $y = \epsilon^{\alpha\beta\gamma\delta} F_{\alpha\beta} F_{\gamma\delta} / \sqrt{-g}$ are

³ Note that the introduction of a metric tensor is just a trick to calculate the symmetric energy-momentum tensor. With some care, the same result can be obtained by canonical methods.

the two possible invariant combinations of field strengths. We get

$$T_{\alpha\beta} = 4F_\alpha{}^\gamma F_{\beta\gamma} \frac{\partial \mathcal{L}}{\partial x} + g_{\alpha\beta} y \frac{\partial \mathcal{L}}{\partial y} - g_{\alpha\beta} \mathcal{L} \quad . \quad (\text{A.3})$$

In the following we will ignore the dependence on y which is irrelevant for the discussion in this appendix. In a purely magnetic field

$$\begin{aligned} F_{0i} &= 0 \\ F_{ij} &= -\epsilon_{ijk} B^k \end{aligned} \quad (\text{A.4})$$

and thus

$$\begin{aligned} T_{00} &= -\mathcal{L} \\ T_{0i} &= 0 \\ T_{ij} &= \delta_{ij} \left(\mathcal{L} - B^k \frac{\partial \mathcal{L}}{\partial B^k} \right) + \frac{\partial \mathcal{L}}{\partial B_i} B_j \quad , \end{aligned} \quad (\text{A.5})$$

which we can compare with our "elementary physics" result from section 3 (slightly generalized),

$$\begin{aligned} T_{00} &= u \\ T_{ij} &= \delta_{ij} \left(-u + B^k \frac{\partial u}{\partial B^k} \right) - \frac{\partial u}{\partial B_i} B_j. \end{aligned} \quad (\text{A.6})$$

Thus the results do agree if $u = -\mathcal{L}$, as indeed it is [2]. We can now proceed to use the field theory calculation to show that the coefficient in the trace anomaly really is the β -function. To this end it is convenient to rescale the gauge fields by $gA_\mu \rightarrow A_\mu$. Then the coupling constant only occurs as a coefficient in front of the gauge theory action,

$$S = \int d^4x \left\{ -\frac{1}{4g^2} F^{\mu\nu} F_{\mu\nu} + \dots \right\} \quad , \quad (\text{A.7})$$

and we can define an effective coupling constant in the renormalized effective action by writing

$$\Gamma = \int d^4x \mathcal{L} = \int d^4x \left[-\frac{1}{4g_{eff}^2} F^{\mu\nu} F_{\mu\nu} \right] \quad . \quad (\text{A.8})$$

This expression defines an effective (or running) coupling constant g_{eff} which is a function of the original renormalized coupling $g(\mu)$, μ and the invariants x and y . Recall (section 2) that μ is the renormalization scale. Since according

to (A.5) $-\mathcal{L}$ is nothing but the renormalized energy density of the vacuum, it is an observable and thus cannot depend on the renormalization point μ .

Combining our equations we get

$$T^\alpha{}_\alpha = \left[4 \frac{x}{g_{eff}} \frac{\partial g_{eff}}{\partial x} \right] \frac{1}{2g_{eff}^2} x \quad . \quad (\text{A.9})$$

Dimensional analysis implies that the μ -dependence of g_{eff} is of the form

$$g_{eff} = g_{eff}\left(\frac{x}{\mu^4}, \frac{y}{\mu^4}, g(\mu)\right) \quad , \quad (\text{A.10})$$

and since \mathcal{L} and thus g_{eff} does not depend on the arbitrary scale μ , we get

$$0 = \mu \frac{dg_{eff}}{d\mu} = -4x \frac{\partial g_{eff}}{\partial x} + \mu \frac{\partial g}{\partial \mu} \frac{\partial g_{eff}}{\partial g} \quad , \quad (\text{A.11})$$

and thus

$$T^\alpha{}_\alpha = \mu \frac{\partial g}{\partial \mu} \frac{\partial g_{eff}}{\partial g} \frac{1}{2g_{eff}^3} F^{\mu\nu} F_{\mu\nu} \quad . \quad (\text{A.12})$$

The β -function is actually the component of a vector field in the space of coupling constants:

$$\beta(g) \frac{\partial}{\partial g} = \mu \frac{\partial g}{\partial \mu} \frac{\partial}{\partial g} \quad . \quad (\text{A.13})$$

It follows that

$$\mu \frac{\partial g}{\partial \mu} \frac{\partial g_{eff}}{\partial g} = \beta(g_{eff}) \quad , \quad (\text{A.14})$$

and thus

$$T^\alpha{}_\alpha = \frac{\beta(g_{eff})}{2g_{eff}^3} F^{\mu\nu} F_{\mu\nu} \quad , \quad (\text{A.15})$$

which, up to the rescaling of the fields and to lowest non-trivial order in perturbation theory, agrees with the result derived in section 3 and moreover makes the relation to the β -function manifest.

The vacuum energy derivation of the trace anomaly given in section 3 makes use of an expression for the energy that is essentially a one-loop result, whereas the arguments in this appendix seems to involve no such approximation. In principle we could calculate the background field effective action to any desired

loop order, and it can still always be written in the form (A.8). For this conclusion we have to keep in mind that there are two ways of presenting the anomaly. One is to establish an equality between the operator T_μ^μ and the renormalized field operators of the theory. The other is to deal directly with matrix elements for T_μ^μ for states where there is an external gauge field. Both the vacuum-energy derivation and the background field effective action derivation follow the second way. It is known that as an operator equality (A.15) is not true in general (beyond one loop). For further discussion, see the review [12] by M. A. Shifman.

Appendix B.

In section 2 we appealed to Lorentz invariance to get a relation between ϵ and μ . Actually it is not quite obvious that we can do so. After all, we want the relation in the presence of an external field so the physical situation is not Lorentz-invariant. We can use the effective action formalism to see under what circumstances we can still use the naive argument. It suffices for our purposes to consider the case of a slowly varying external abelian color electromagnetic field. Then again $\mathcal{L} = \mathcal{L}(x, y^2)$ with $x = 2(B^2 - E^2)$ and $y = 4\mathbf{E} \cdot \mathbf{B}$. (Since y is a pseudoscalar it can't occur in odd powers.) We get [13]

$$\begin{aligned} \mathbf{D} &= \frac{\partial \mathcal{L}}{\partial \mathbf{E}} = -4 \frac{\partial \mathcal{L}}{\partial x} \mathbf{E} + 8y \frac{\partial \mathcal{L}}{\partial y^2} \mathbf{B}, \\ \mathbf{H} &= -\frac{\partial \mathcal{L}}{\partial \mathbf{B}} = -4 \frac{\partial \mathcal{L}}{\partial x} \mathbf{B} - 8y \frac{\partial \mathcal{L}}{\partial y^2} \mathbf{E} \quad . \end{aligned} \quad (\text{B.1})$$

We see that if $\mathbf{E} \cdot \mathbf{B} = 0$ and if we define

$$\mathbf{D} = \epsilon \mathbf{E}, \quad \mathbf{B} = \mu \mathbf{H} \quad , \quad (\text{B.2})$$

we do get

$$\epsilon \mu = 1 \quad , \quad (\text{B.3})$$

(assuming the derivatives do not become singular as $y \rightarrow 0$). In particular, we can use this relation for a homogeneous magnetostatic field.

References

- [1] R. Peierls, quoted in L. Hoddeson, G. Baym and M. Eckert, *Rev. Mod. Phys.* **59** (1987) 289

- [2] V.S. Weisskopf, *Mat. Fys. Medd. Dan. Vid. Selsk.* **XIV** no 6 (1936).
- [3] N.K. Nielsen, *Am. J. Phys.* **48** (1981) 1171.
- [4] R.J. Hughes, *Nucl. Phys.* **B186** (1981) 376.
- [5] G. 't Hooft, quoted in P.D. Mannheim, *Phys. Rev.* **D 9** (1974) 3438.
- [6] V.B. Berestetskii, E.M. Lifshitz and L.P. Pitaevskii, *Quantum Electrodynamics*, §114, Pergamon Press, London (1982).
- [7] L. F. Abbott, *Nucl. Phys.* **B185** (1981) 189.
- [8] J.D. Jackson, *Classical Electrodynamics* (2nd ed.), Wiley, New York (1975).
- [9] L.D. Landau, E.M. Lifshitz and L.P. Pitaevskii, *Electrodynamics of Continuous Media* (2nd ed.), §15, Pergamon Press, London (1984).
- [10] J. Polchinski, *Nucl. Phys.* **B303** (1988) 226.
- [11] H. Pagels and E. Tomboulis, *Nucl. Phys.* **B143** (1978) 485.
- [12] M. A. Shifman, *Phys. Rep.* **209** (1991) 341.
- [13] V.B. Berestetskii, E.M. Lifshitz and L.P. Pitaevskii, *Quantum Electrodynamics*, §130, Pergamon Press, London (1982).